

Three speakers, four dialects: Documenting variation in an endangered Amazonian language

Amalia Skilton
University of California, Berkeley

This paper offers a case study on dialect contact in Máihiki (Tukanoan, Peru), with the goal of illustrating how documentation of variation can contribute to a general language documentation project. I begin by describing the facts of variation in one dialectally diverse Máihiki-speaking community. I then argue that the outcomes of dialect mixing in this speech community can be understood only through a fine-grained analysis centering the dialectal composition of the communities of practice to which speakers belonged in early life. The coarse-grained identity categories used in most variationist analyses, such as age and gender, are less informative. After proposing a network theory interpretation of this finding, I discuss its implications for the role of (a) ethnography and (b) the European dialect mixing literature in research on variation in endangered languages. Second, I describe some surprising similarities between this speech community and those described in classic variationist literature. Like urban English speakers, Máihiki speakers attach less indexical value to morphosyntactic than to phonological variation, and – although their language lacks a standard – engage in indexically motivated style-shifting. I discuss ways to adapt variationist methods to endangered language settings to capture these phenomena, then close with comments on the importance of documenting variation for conservation.

1. INTRODUCTION. This paper aims to illustrate some of the contributions which documentation of inter-speaker variation can make to a general language documentation project.¹ Drawing on my own experience conducting fieldwork on variation in a critically endangered language, I argue here that studying variation is valuable – and in some cases, indispensable – to achieving the most common goals of language documentation projects. Where documentation materials will be used for language conservation activities, documentation of variation is necessary for the materials to accurately represent the internal diversity of the language. Likewise, where documentation will be the basis for

¹ I owe endless thanks to all of the Máihuna people who have shared their language with me, especially Federico, Soraida, and Pedro López Algoba; Otilia López Gordillo; Trujillo Ríos Díaz; and Adriano Ríos Sanchez: *misàrè góásàgànikò bàidzi*. For feedback, I thank Lev Michael; Christine Beier; Stephanie Farmer; audiences in Honolulu and Berkeley; the other participants in the “Documenting Variation” session, especially Sali Tagliamonte and the editors of the volume; and an anonymous reviewer from *LD&C*. I also thank Leanne Hinton and Justin Spence for thoughts relevant to the conclusion, and the Vicariato San José del Amazonas for logistical support during fieldwork. This research was supported by a Parker Huang Undergraduate Research Fellowship from Yale University, an NSF Graduate Research Fellowship, and NSF BCS-1065621. All errors and omissions are my own.

grammatical description, variation can be crucial in weighing diachronic and synchronic explanations for generalizations (Rice 2006: 402-403).

Yet, for several reasons, general documentation projects do not examine variation as often as they might (though see Rice 1989, ch. 4, and Campbell 2014, ch. 6, for two exceptions). It is unusual for students to be trained in both cross-linguistic descriptive fieldwork and sociolinguistics or dialectology. Field methods textbooks (e.g. Samarin 1967, Vaux & Cooper 1999, Bowerman 2008) and language documentation literature (e.g. Austin & Sallabank 2011, Chelliah & DeReuse 2011) say relatively little about how to investigate and describe variation. And a survey of reference grammars suggests that many descriptive and documentary projects take much of their data from a small group of speakers who display only modest variation.

Moreover, research on variation in endangered and poorly described languages faces particular obstacles. It is difficult to record casual speech, showing the full range of variation, from rusty or self-conscious speakers. Sociolinguistic variation in a non-industrial society may not follow the same social patterns that are highlighted in the sociolinguistic literature on the US and Europe. The linguist may be unfamiliar with the sociolinguistic variables, or may not expect variation in some areas of the grammar. Or, especially in the early stages of fieldwork, language ideologies may conceal variation from the linguist entirely. And while sociolinguists working in industrial societies have devised methods to address some of these problems, many common sociolinguistic research methods – for example, interview modules which involve reading passages or wordlists – are inappropriate in endangered language settings.

Sociolinguistics has thus continued to focus primarily on varieties of English and other Indo-European languages, even as other subfields of linguistics have taken a strong cross-linguistic turn. As a consequence of this focus, the cross-linguistic and cross-cultural applicability of many sociolinguistic (and therefore also historical linguistic) generalizations remains to be tested (Stanford & Preston 2009; Stanford 2008, 2009).

Against this background, this paper describes the methods which I used to carry out sociolinguistic documentation of Máihiki, a severely endangered Amazonian language, in a field situation with massive variation produced by dialect mixing. My goals are two. First, I sketch some of the differences in the structure of variation between one Máihiki-speaking community and some speech communities described in well-known sociolinguistic literature (such as Trudgill 1974, Milroy & Milroy 1985, Eckert 1988, or the studies summarized in Labov 2001). Like many communities where language documentarians work, the speech community which I discuss is small, non-industrial, and has been displaced from its traditional lands. My description of variation in this setting will therefore, I hope, give other documentarians a sense of where – in which linguistic features, and among which speakers – they may encounter variation at their own field sites. Second, I illustrate how general language documentation projects can incorporate documentation of variation, including by adopting methods from variationist sociolinguistics.

The paper is organized as follows. The following section (§2) describes the speech community and language of study. §3 gives a case study of four sociolinguistic variables in this speech community, first summarizing the facts of variation (§3.1) and then arguing for an ethnographically grounded analysis of the social patterning of the variation (§3.2, §3.3). §4 discusses ways in which variation in Máihiki departs, or

interestingly fails to depart, from the kinds of variation discussed in the sociolinguistic literature cited above. These include the dominance of morphosyntactic over phonological variation (§4.1) and the presence of indexically motivated style-shifting (§4.2). §5 presents the methods which I used to integrate collection of data on and analysis of this variation into a general documentation project. In §6, I summarize and argue for the importance of studying variation to both language documentation and conservation generally.

2. BACKGROUND: THE LANGUAGE AND FIELD SITE. Máihiki (Western Tukanoan, ISO-639: ore; also known by the pejorative terms ‘Orejón’ and ‘Coto’) is the traditional language of the Máihuna ethnic group, an indigenous people of northwestern Amazonia. As shown in Figure 1, Máihuna people traditionally occupied the area between the Napo and Putumayo Rivers, located in the extreme northeast of contemporary Peru. They subsisted by manioc horticulture, wild-gathering, hunting, and fishing. Unlike some well-known Tukanoan peoples of the Vaupés area (Sorensen 1967, Hugh-Jones 1979, Jackson 1983), and like other Western Tukanoan peoples, Máihuna people traditionally practiced clan exogamy and language group endogamy, and did not stigmatize code-switching or lexical borrowing.

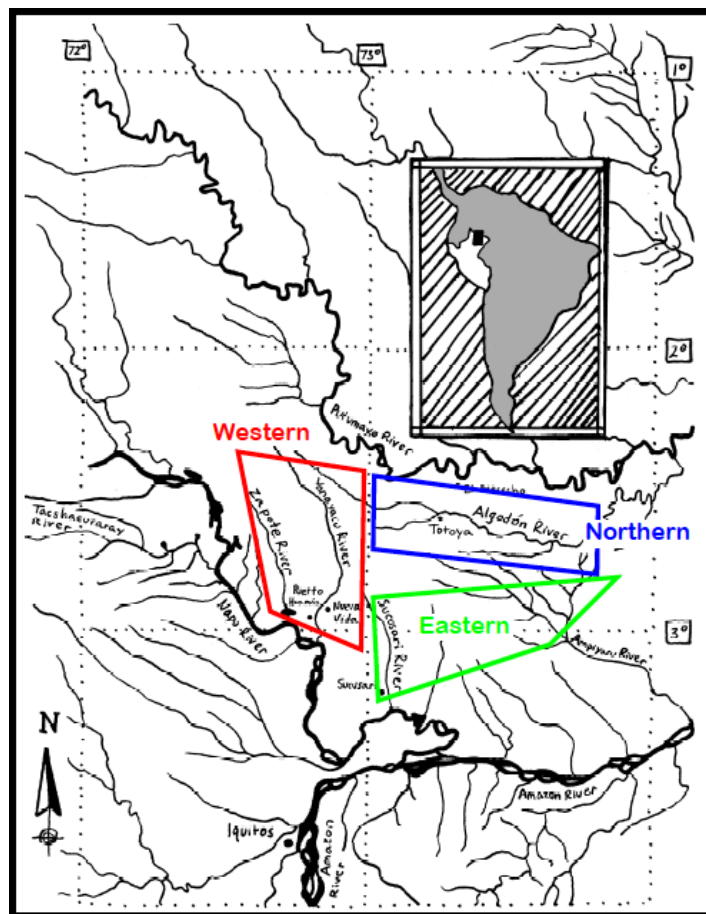


FIGURE 1: Map of Máihuna settlement areas by dialect. Inset shows location of Máihuna area within Peru. Base map by Farmer (2015: 5).

The Máihiki language is severely endangered, with ~75 speakers in the ethnic Máihuna population of ~500. The speakers, most of whom were born before 1965, are scattered across five widely separated settlements, all but one within the traditional territory. All Máihiki speakers speak at least some regional Spanish, and for most, Spanish is the dominant language of daily interaction. No children have fully acquired Máihiki since the late 1970s.

Máihuna people distinguish three mutually intelligible dialects of Máihiki, which they typically describe in terms of lexical differences. While speakers tend to refer to dialects by the names of the river basins where they are spoken, I refer to them here as ‘Northern,’ ‘Eastern,’ and ‘Western.’ Figure 1 shows the territory traditionally associated with each dialect. In 2014, the Northern dialect, associated with the Algodón River (a tributary of the Putumayo River), had 12 fluent speakers. The Eastern dialect, associated with the Sucusari River (a tributary of the Napo River), had 15 speakers. All of the other ~50 speakers of Máihiki spoke varieties of the Western dialect, associated with the Yanayacu River (a tributary of the Napo). Historically, there was a fourth dialect, closely related to the Eastern dialect, associated with the Ampiyacu and Apayacu rivers (tributaries of the Amazon). This variety has never been documented and may no longer be spoken.

I conducted the fieldwork described here between January and June 2014 in San Antonio del Estrecho, a multi-ethnic town of ~3,000 people located on the Putumayo River (a tributary of the Amazon), which forms the border between Peru and Colombia. Estrecho, as it is called, is one day's walk overland from the traditional Máihuna settlement area on the Algodón River. At the time of my research, ~70 Máihuna people, representing at least three-quarters of the total Máihuna population of this region, had left the Algodón and come to live in Estrecho. Their migration formed part of a long, colonially mediated process in which Máihuna people in the Putumayo basin have shifted from living in mobile bands in remote interfluvial areas, to living in fixed settlements on or near the main course of the Putumayo (Skilton 2014: 4-6.)

My goal in working with the former Algodón speech community was to produce a grammatical sketch of Northern Máihiki for a polylectal descriptive grammar, since the Northern dialect was the most endangered and least described of the varieties of Máihiki.² Local speakers, aware of differences between their speech and that of other dialect areas, supported this goal. Yet within a few weeks of my arrival, it became clear that there was no invariant ‘Northern dialect’ to take as the object of grammatical description.

3. THE SOCIAL PARAMETERS OF VARIATION. Máihiki speakers born in the Algodón form a small and tight-knit speech community, but exhibit exuberant variation. Speakers who are closely related and have been coresident for most of their lives, such as siblings, display different reflexes of many sociolinguistic variables.

In this section, I summarize the facts of phonological and morphological variation in Northern Máihiki (§3.1) and describe the social situation which gave rise to this variation (§3.2). My account of the patterning of variation focuses on the life histories

² SIL members published short descriptions of the Eastern dialect in the 1960s and 1970s (Velie 1975, Velie & Velie 1981). My collaborators Lev Michael, Christine Beier, and Stephanie Farmer had carried out several months of fieldwork on the Western dialect (Michael et al. 2014; Farmer 2015).

and kin relations of speakers, rather than on their memberships in corporate groups³ such as age cohorts, genders, or clans. These more familiar sociolinguistic factors, I show in §3.3, capture almost none of the variation found in Northern Máihiki. By this, I do *not* mean that different corporate groups are relevant in this setting – for instance, that socioeconomic status operates differently there (cf. Rickford 1986) or that clan is important and class is not (Stanford 2009). Rather, I aim to show that corporate group memberships *in general* are less relevant, and the details of life history more relevant, to analysis of variation in Máihiki.

3.1 The facts of variation. Table 1 shows the values of four sociolinguistic variables for six Máihiki speakers born in the Algodón basin, a place understood by speakers as forming a single settlement area. Two variables, (VHV) and (K^wA), are phonological – they represent sound changes which, at some point, affected the entire lexicon. The other two, (3SG.PST) and (go.1SG.PST), are morphological – they represent analogy-driven change in the form of specific morphological constructions.⁴ Both the phonological and the morphological variables are categorical, in the sense that I never observed forms intermediate between the variants shown.⁵

	(VHV): 'descend'	(K ^w A): 'cook.VT'	(3SG.PST.NF/F)	(go.1SG.PST)
Soraida	gàè	kòkò	-ai/-ao	sá-hì
Pedro	gàè, gáhè	kòkò (text), k ^w àkò (list)	-ai/-ao	sái-bì
Federico	gàè, gáhè	kòkò (text), k ^w àkò (list)	-ai/-ao	sái-bì
Trujillo	gàè, gáhè	kòkò (text), k ^w àkò (list)	-gi/-go	sá-hĩ
Adriano	gàè	kòkò	-gi/-go	sá-hĩ
Otilia	gàè	kòkò	-ai/-ao	sàà-bì

TABLE 1: Four sociolinguistic variables in Northern Máihiki.

³ A 'corporate group' is a group of people that shares rights and duties, is relatively stable in membership, and outlasts the deaths or separation from the group of its members. For example, nuclear families, socioeconomic classes, and clans are all potentially corporate groups. Hayden & Cannon (1982: 133-134) discuss this term and its history in more detail.

⁴ Glossing in this paper follows Leipzig conventions (except that NF is 'non-feminine'): 1 'first person', 2 'second person', 3 'third person', E 'Eastern', F 'feminine', N 'Northern', NF 'nonfeminine', PL 'plural' PST 'past', SG 'singular', VT 'transitive verb', W 'Western'.

⁵ §3.2.2 provides complete linguistic descriptions of the variables. In brief: the variable (VHV) represents the reflexes of roots with Pre-Máihiki (PM) intervocalic *h, exemplified by the verb *gáhè 'descend, go downriver.' (K^wA) represents reflexes of roots with PM initial *k^w, exemplified by *k^wako 'cook.VT'. Máihiki has two verb classes: a large open class, and the 'i-class,' a closed class which follows a different inflectional paradigm and displays root allomorphy governed by the first affix right of the root. (3SG.PST) represents the affixes which, on open-class verbs, mark past tense and 3SG subject agreement (forms on the left of the slash agree with non-feminine subjects, forms on the right with feminine subjects). (go.1SG.PST) represents the morphological construction, consisting of a root allomorph and agreement affix, used to mark past tense with (a) 1SG and (b) all plural subject agreement on i-class verbs, such as 'go.'

Table 1 combines text and wordlist data (cf. §4). While there is intraspeaker variation in the phonological variables (VHV) and (K^wA), the data contains no evidence of intraspeaker variation in the morphological variables (3SG.PST) and (go.1SG.PST).

The data in Table 1 exemplifies two general attributes of variation in this speech community: (a) the variants are numerous, and (b) for some variables, variants cannot be categorized into isogloss bundles even at the level of the idiolect. On (a), the six speakers in Table 1 display four different forms for the variable (go.1SG.PST). One of the four variants, *sáhi* (Soraida), is identical to the form found in Western and Eastern Máihiki. Another variant, *sáhĩ* (Trujillo, Adriano) reflects a minor phonological change from this form – nasalization of the final vowel, likely due to rhinoglottophilia. Two variants – *sáibi* (Pedro, Federico) and *sáàbi* (Otilia) – represent significant morphological innovations relative to the forms with *h*. On (b), there are only two speakers in Table 1, Pedro and Federico, who display exactly the same variants. Other pairs, such as Adriano and Soraida, or Trujillo and Pedro, display the same phonological variants, but different morphological ones.

The extent of the variation shown in Table 1 is more surprising because the six speakers whose data is shown form a single, close-knit social network. At the time of my research, the speakers listed in Table 1 live in the same location, as they have for most of their lives, and communicate with each other in Máihiki every day, though some pairs interact in Spanish as well. In addition, all of the people listed in Table 1 are close kin. Every speaker shown has a first- or second-degree kinship relation with every other one, and the list includes two sets of siblings (cf. Figure 2 below). In the terms of network theory as operationalized for sociolinguistics by Gal (1979) and Milroy (1980), all of the speakers in Table 1 have strong, multiplex ties – by kinship and by coresidence – with one another, and they have relatively few ties to fluent Máihiki speakers outside the group. Small, close-knit networks have been claimed to suppress variation (Milroy & Milroy 1985: 364, Lippi-Green 1989: 217–218). But here, variation flourishes among the Table 1 speakers despite the strong and multiplex ties between them.

3.2 The source of variation: dialect contact. The variation shown in Table 1 is the consequence of almost a century of intensive dialect contact between Northern, Eastern, and Western Máihiki in the Algodón basin, the traditional Northern Máihiki dialect area. Máihuna oral history indicates that Máihuna people from the Eastern and Western dialect areas began moving to the Algodón in the 1920s, possibly brought by labor patrons. Migration from other areas to the Algodón accelerated in the 1960s, when a large number of young men from the Algodón began traveling to Eastern and Western Máihiki-speaking communities to find wives, then bringing their brides – and, in many cases, the brides' families – back to the Algodón with them. (This 'spouse import' was driven by severe demographic imbalances in the Algodón population. It is not a traditional Máihuna practice.)

Influential studies of koineization in European languages by Kerswill & Williams (2000), Trudgill et al. (2000), Trudgill (2004), and Kerswill & Trudgill (2005) have suggested that, in situations of intense contact between speakers of mutually intelligible dialects, dialect differences will be leveled – leading to the emergence of a uniform koiné – within one to two generations. Yet dialect contact in the social network which I describe here did not lead to the emergence of a koiné, even though it preceded language

shift by at least five decades. Rather, each Máihiki speaker born in the Algodón during the dialect contact period displays a different mixture of Northern, Western, and Eastern dialect features.

In this section, I begin by tracing the kinship and coresidence relations of the speakers shown in Table 1 (§3.2.1). I then suggest, in §3.2.2, that the outcome of dialect mixing for each individual in this population directly reflects the dialectal composition of the gender-, kinship-, and coresidence-based communities of practice (Lave & Wenger 1991) to which he or she belonged in early life. In §3.3, I contrast the explanatory power of this very fine-grained, ethnographic analysis with the power of an analysis that focuses on the kind of coarse-grained identity categories, such as age and gender, that are often prominent in variationist analyses. My evaluations of the analyses in §3.2 and §3.3 are exclusively qualitative, since at the time of my research, only 12 Máihiki speakers born in the Algodón were still living – a population too small for statistical analysis to be meaningful.

3.2.1 Kinship and place. Gender, age, and clan membership are central organizing principles in Máihuna society and traditional ritual life (Bellier 1991). Yet when the speakers in Table 1 talk about the ways that they are different from one another and from other Máihuna people, they do not treat these categories as most salient. What they see as relevant, instead, are the places with which a person and their kin are associated: where someone was born, where they lived as a child, and where their parents were born. These categories assume central importance in my analysis of variation as well, for it is kinship and place – as indices of the communities of practice to which speakers have belonged – which best explain the variation shown in Table 1.

I therefore now summarize the kinship and coresidence relations among the speakers listed in Table 1. Figure 2 displays a genealogical diagram of the Table 1 speakers, overlaid with information about place. Each shape in Figure 2 represents a named individual who is listed in Table 1 or is the parent or spouse of a speaker listed there. Information in Figure 2 was collected in genealogical and life-history interviews with the speakers listed in Table 1 and confirmed, where possible, with missionary records. Squares represent men, circles women.

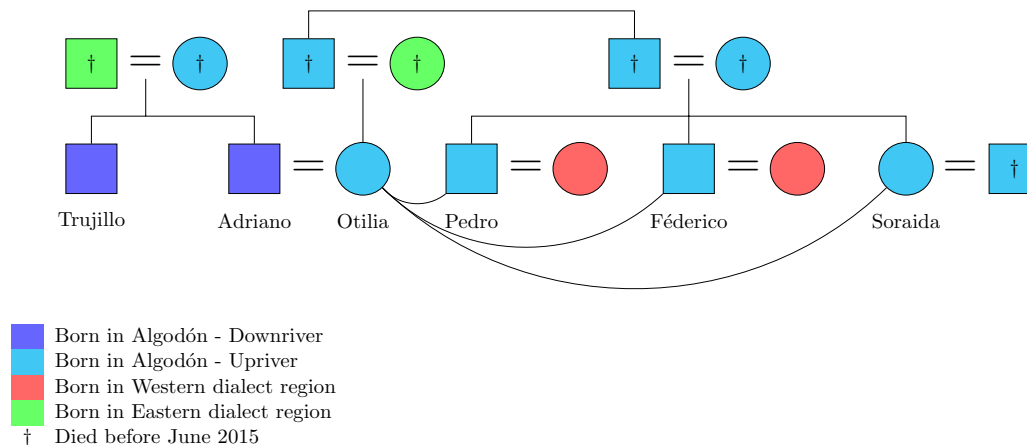


FIGURE 2: Kin relations among speakers in Table 1.

Figure 2 shows that Adriano and Trujillo are siblings, as are Soraida, Pedro, and Federico. Otilia is married to Adriano and is in a consanguineal relation to Soraida and her brothers through which she is classified as their sister. Because of this classificatory relationship, represented by the curved line, Otilia links the two sets of biological siblings into a single kin network. Soraida and her siblings are entitled to address Otilia as *dòikò* ‘sister,’ her husband Adriano as *bât* ‘brother-in-law,’ and her children as *hĩtĩ* ‘nephew’ and *hĩtò* ‘niece.’

The contrast between blue and cyan shapes in Figure 2 represents historical coresidence relations among the speakers born in the Algodón. Six people here – Soraida and her brothers, whose variants are shown in Table 1, and three deceased people – were born far upriver on the Algodón. They lived in extremely remote locations in the headwaters of the Algodón and in the area between the Algodón and Napo until the early 1950s, when missionaries persuaded them to settle in a village on the more accessible middle Algodón. Otilia was born at this new settlement shortly after it was founded. Two other people on the diagram – Adriano and Trujillo, both also in Table 1 – were born in an extended-family settlement located downriver on Algodón, near the river’s mouth. This settlement, founded by Adriano and Trujillo’s father in the 1920s, dissolved in the mid-1950s. Its members then migrated to the village on the middle Algodón.

Finally, the contrast between green, red, and blue or cyan shapes on Figure 2 visually represents the birthplace of each person. Except for Soraida, every Table 1 speaker has at least one first-degree relative – parent or spouse – born outside the Algodón. Adriano and Trujillo’s father was born in the Sucusari basin, where Eastern Máihĩki is currently spoken. Otilia’s mother was born in the Ampiyacu-Apayacu basin, which is geographically very close to the Sucusari basin, and likely spoke a dialect similar to Eastern Máihĩki. Pedro and Federico are married to Western Máihĩki-speaking women from the Yanayacu basin.

3.2.2 Communities of practice. In Table 2, almost all of the variation shown can be understood by reference to the three social factors displayed in the three left columns: (a) the speaker's place of birth and childhood residence, (b) the dialects spoken by his or her parents (inferred from the parents' birthplaces, as all of the speakers' parents are deceased), and (c) the dialect spoken by his or her spouse. Table 2 re-displays a version of Table 1 expanded to include this information, as well as the forms found in the Eastern and Western dialects.

	Birth Group	Parent Dial.	Spouse Dial.	(VHV): 'descend'	(K ^w A): 'cook.VT'	(3SG. PST)	(go.1SG. PST)
Soraida	Upriver	N, N	N	gáè-	kòkò-	-ai/-ao	sáhì
Pedro	Upriver	N, N	W	gáè-, gáhè-	kòkò-, k ^w àkò-	-ai/-ao	sáíbì
Féderico	Upriver	N, N	W	gáè-, gáhè-	kòkò-, k ^w àkò-	-ai/-ao	sáíbì
Trujillo	Downriver	E (Sucu-sari), N	N/A	gáè-, gáhè-	kòkò-, k ^w àkò-	-gi/-go	sáhĩ
Adriano	Downriver	N ⁶	N	gáè-	kòkò-	-gi/-go	sáhĩ
Otilia	Upriver	E (Ampi-yacu), N	N	gáè-	kòkò-	-ai/-ao	sáábì
Eastern dial. (Sucusari)	N/A	N/A	N/A	áhè-	k ^w àkò-	-gi/-go	sáhì
Western dial.	N/A	N/A	N/A	gáhè-	k ^w àkò-	-gi/-go	sáhì

TABLE 2: Four sociolinguistic variables and three parameters of variation

(3SG.PST). All finite verbs in Máihĩki bear inflectional suffixes which fusionally expone tense, sentential mood, and the subject features of person, number, and gender (feminine vs. non-feminine). This sociolinguistic variable relates to the segmental form of the affixes which mark past tense and third person singular agreement on the main class of verbs. In Pre-Máihĩki (PM), the third person singular non-feminine past tense (3SG.NF.PST) affix reconstructs to **-agi*, and its feminine (3SG.F.PST) counterpart to **-ago*. In the contemporary language, there are two reflexes of these forms: *-ai* (NF)/*-ao* (F), used only by speakers born in the Algodón, and *-gi* (NF)/*-go* (F), which is used by some speakers in the Algodón and all speakers in the other two dialect regions.

Among the Table 1 speakers, place of birth within the Algodón basin has a clear relationship with this variable. Trujillo and Adriano, the two speakers born in the downriver settlement, are the only speakers here to use the *-gi/-go* variants. These variants are also found in Eastern Máihĩki, the dialect spoken by their late father, the founder of that settlement. All other speakers in Table 1 (and to my knowledge, all speakers born in the Algodón except for the former members of the downriver

⁶ Adriano and Trujillo, though full siblings, are assigned different values of Parent Dial because their father died when Adriano was approximately two years old and Trujillo, who is at least ten years older, was in his teens.

settlement) use the *-ai* variant, which is found exclusively in the Northern dialect region. Trujillo and Adriano's use of *-gi* therefore likely reflects that their input in acquisition included more Eastern Máihiki and Eastern-influenced varieties – that is, their father's speech and the speech of their older siblings, who would have been the brothers' main caregivers – than Northern Máihiki. The brothers' variant of (go.1SG.PST), *sáhĩ*, also closely resembles the form found in Eastern Máihiki.

This does not mean, though, that Trujillo and Adriano's speech is Eastern Máihiki. In addition to the morphological variables, the Eastern dialect as spoken in its historical territory also displays regular changes of PM root-initial **g > Ø* and root-initial **g^w > b*. Adriano and Trujillo's speech never displays **g > Ø*, and it displays **g^w > b* in only one word. Instead, the brothers display – in all styles for Adriano, and in connected speech for Trujillo – the sound changes characteristic of Northern Máihiki, root-medial PM **h > Ø* and initial **k^w > k*, discussed below. These phonological variants likely reflect that the brothers' language, and that of their older siblings, was influenced by the Northern Máihiki speech of their mother and maternal relatives, who resided with them in the downriver settlement. It is unlikely that Trujillo and Adriano's Northern variants are due to accommodation to Northern dialect speakers, since – among other members of their community – even extremely long-term accommodation has not led to change in the variants used in connected speech (cf. Labov 1981, which shows that even child migrants to Philadelphia do not fully acquire the Philadelphia short /a/ system).

(go.1SG.PST). Like all Western Tukanoan languages, Máihiki displays two classes of verbs: a large open class, and the small, closed 'i-class.' I-class verbs follow a different inflectional paradigm from the open class. They also exhibit root allomorphy governed by the suffix immediately following the root, varying between allomorphs of the form CVi (e.g. *sái*), CV (*sá*), CVV (*sáà*), and CVni (*sáni*). This variable represents the morphological construction – the root allomorph and inflectional suffix – used to mark past tense and (a) 1SG and (b) all plural subject agreement on i-class verbs, which are exemplified here by the verb 'go.' In PM, this construction reconstructs as **sá-hĩ*, with the i-class root appearing in its CV allomorph and bearing the inflectional suffix *-hĩ*. All Máihiki speakers outside the Algodón basin either preserve the PM form of this variable, *sá-hĩ*, or display a variant, *sá-hĩ̃*, which differs from the PM form only in that the final vowel has nasalized. In the Algodón, there are two more reflexes of (go.1SG.PST): *sáà-bĩ*, a form which reflects extension of the open-class inflection to the i-class, and *sái-bĩ*, which reflects extension of the open-class inflection followed by leveling of the i-class root allomorphy.

The social distribution of the reflexes of (go.1SG.PST) is explained by the intersection of place of residence and gender. The four speakers born in the upriver group display three forms for this variable: Soraida and Otilia show *sáhĩ* and *sáàbĩ*, while the two men, Pedro and Federico, use *sáibĩ*. They share this variant, which is innovative relative to *sáhĩ* and *sáhĩ̃*, with an older, genealogically unrelated man from the upriver group (not shown in the table), Jorge.

How did Pedro and Federico come to use *sáibĩ* when Soraida, their sister, uses *sáhĩ*? The answer cannot lie in the earliest stages of acquisition – the same caregivers cared for all three siblings as young children, and all three would have encountered near-identical input in early acquisition. The variation also does not represent absolute gender indexicality, since speakers do not shift between the forms in reporting one another's

speech. It is more probable that Soraida and her brothers came to employ these forms because they participated in distinct, gendered communities of everyday practice within the upriver group. The division of labor by gender was strict in traditional Máihuna society (Bellier 1991: 76-120), and most interaction between unrelated people still takes place in single-gender groups. Pedro and Federico likely settled on the *sáibì* variant because, as children and young men, they co-participated with older men of their group, such as Jorge, in exclusively male subsistence and extractive activities. Soraida's more conservative variant, *sáhì*, may have been associated with a separate community of practice organized around women's activities, or may have been the majority variant at this time. It is unlikely that she acquired it by accommodation to Western or Eastern Máihiki speakers, since she otherwise displays no discernible Eastern or Western influence. Otilia's form, *sààbì*, was almost certainly acquired from her mother, a speaker of the Ampiyacu-Apayacu variety of Eastern Máihiki.

Phonological variables. Three of the six speakers in Table 1 display regular sound changes of PM root-medial **h* > Ø and initial **k^w* > *k*, exemplified by their forms of 'descend' and 'cook.' These sound changes are not found outside the Algodón, and forms with the changes are metalinguistically treated by speakers as indexical of Algodón speech. The other three speakers in Table 1 – Trujillo, Pedro, and Federico – used the innovative variants of (VHV) and (K^wA) in casual speech, but gave more conservative variants retaining **h* and **k^w* in wordlist elicitation and extremely careful speech. This is the canonical profile of a sound change in progress, but here it is more likely to reflect a completed sound change partially undone by dialect contact. Trujillo, Pedro, and Federico are the speakers who have had the most lifetime contact with Western dialect speakers, for Pedro and Federico are married to Western dialect-speaking women, and Trujillo (a bachelor) has resided for most of his adult life in a household headed by one of his younger brothers and the brother's Western dialect-speaking wife.

Like most Western dialect speakers living in the Northern dialect region, Pedro and Federico's wives explicitly judge their own conservative variants of (VHV) and (K^wA) 'more correct' than the innovative variants.⁷ They often volunteered metalinguistic comments to me about the 'incorrect' status of forms like *kòkò*, the innovative variant of 'cook,' and their Northern dialect-speaking husbands gave the same indexical judgment when asked to reflect on the contrast between *k^wàkò* and *kòkò*. Although Northern dialect speakers who do not have Western dialect-speaking intimates are aware of the difference between their speech and the Western dialect in reflexes of these items, they do not view their own reflexes as 'incorrect.' Trujillo, Pedro, and Federico's variation between Northern and Western variants of (VHV) and (K^wA), therefore, likely reflects change both in their indexical evaluations of these forms and in their usage as a result of contact with Western dialect speakers.

Summary. The data in Table 1 includes reflexes from four dialects in total: the Western dialect, which is the source of some of Pedro, Federico, and Trujillo's phonological variants; the pre-contact form of the Northern dialect, which is the source

⁷ Western dialect speakers in general, not only these women, consider their speech 'more correct' than other dialects of Máihiki. When I visited the Western dialect area in May 2016, Western Máihiki speakers instantly noticed and corrected my use of forms reflecting **h* > Ø and **k^w* > *k*, saying that lexical items such as *ʃí* 'husband' (< *ʃhí*) and *kòkò* 'cook' (< *k^wàkò*) were incorrect. One Western dialect speaker added, speaking of his late Northern dialect-speaking wife: *dʒi kímà hìkàkìrè, kámà dʒétégó jíò*, 'how I talk, she learned (to speak) like that.'

of some phonological variants for all speakers, and of all morphology for Pedro, Federico, and Soraida; the Eastern dialect, which is the source of Adriano and Trujillo's morphology; and the Ampiyacu-Apayacu variety of the Eastern dialect, the likely source of one of Otilia's morphological variants.

Though I have illustrated this variation with data from six speakers, it would have been possible to find reflexes from all four dialects in data from as few as three speakers – for example, in the data from Adriano, his brother Trujillo, and his wife Otilia. Such are the consequences of intense, prolonged dialect contact, in the English-speaking world (Trudgill et al. 2000) as well as the Algodón basin.

3.3 Corporate groups. In §3.2.2, I examined the variation shown in Table 1 by discussing when and how each speaker acquired the variants they display. In this section, I assess the qualitative difference in explanatory power between the ethnographic view proposed in §3.2.2 and an analysis centering on (locally relevant) corporate groups.

As mentioned in §3.2.2, Máihuna people overtly classify one another into three kinds of corporate groups: by gender, by named age categories, and into three exogamous patrilineal clans. Socioeconomic class is not a meaningful variable in this economically flat society, while clans are a central part of social organization, important not only for marriage, but also in politics, classificatory kinship, and (in the past) ritual life. Table 3 re-displays a version of Table 1 expanded to include the gender, birth year, and clan affiliation of each speaker.

	Gender	BirthYr	Clan	(VHV)	(K ^w A)	(3SG.M. PST)	(go.1SG. PST)
Soraida	F	c.1945	ódzò	gàè-	kòkò-	-aì	sáhì
Pedro	M	c.1950	ódzò	gàè-, gáhè-	kòkò-, k ^w àkò-	-aì	sáìbì
Federico	M	c.1955	ódzò	gàè-, gáhè-	kòkò-, k ^w àkò-	-aì	sáìbì
Trujillo	M	c.1930	ídzé	gàè-, gáhè-	kòkò-, k ^w àkò-	-gì	sáhĩ
Adriano	M	c.1945	ídzé	gàè-	kòkò-	-gì	sáhĩ
Otilia	F	1962	ódzò	gàè-	kòkò-	-aì	sààbì

TABLE 3: Four sociolinguistic variables and three corporate groups

Table 3 shows that age, gender, and clan – alone or combined – do not predict variation effectively in this setting. Gender under-predicts the variation: the male speakers differ on all four variables, although the female speakers share variants for three of the four variables. Age is also not informative, since each of the non-sibling pairs nearest in age (Soraida and Adriano, Pedro and Adriano, Federico and Otilia) display different values for at least two variables. Clan initially appears to be relevant: while the *ódzò* clan speakers display wide variation on all variables but (3SG.M.PST), only the two *ídzé* clan speakers display the *-gì* variant for (3SG.M.PST) and the *sáhĩ* variant for (go.1SG.PST). Yet these variants are not associated with the *ídzé* clan in general. In the Western and Eastern Máihĩki dialect areas, people of all clans use the *-gì* variant of

(3SG.M.PST) and the *sáhì* and *sáhî* variant of (go.1SG.PST), which here is associated only with a single *ódzò* speaker. The apparent effect of clan, then, is epiphenomenal on the fact that the *ídzé* people in this table have been more influenced by Eastern Máihîki than the *ódzò* ones.

Coarse-grained social variables like the ones shown in Table 3 are important predictors of variation in many non-industrial societies (see, for instance, Romero 2009 for gender; Clarke 2009 for age; and Garde 2008 and Stanford 2008, 2009 for clan), but they cannot explain the variation found in Northern Máihîki. It is possible that the lack of relationship between variation and corporate group memberships in Northern Máihîki is exceptional, or that it reflects the relatively shallow time depth of the dialect contact. A more likely explanation, though, is that – for people who live in a very-small-scale society and have correspondingly small social networks – each social tie has greater influence on acquisition and usage than in large-scale societies (cf. Givón and Young's 2002 'societies of intimates,' cited in Trudgill 2011: 169). For example, a child who (as two of the speakers in this study did) lives in a settlement some hours' travel from the nearest neighbor, does not attend school, and does not regularly interact with people outside his extended family, is more likely to adopt the variants of his older relatives than to adopt those of agetates who he sees only on occasional visits. Few speakers lead linguistic lives this isolated in industrial societies or in large-scale agrarian societies. In indigenous Amazonian societies inhabiting interfluvial areas, on the other hand, the residence and interaction pattern that I have just described is common (Harner 1972, Johnson 2003).

Recent variationist research, such as Meyerhoff (2015), Stanford and Pan (2013), and the essays in Stanford and Preston (2009), has expanded sociolinguistic knowledge about non-industrial societies. Yet the above suggests that sociolinguistics may still have too little data about such settings to generate strong hypotheses about which aspects of social structure will organize variation there. Sociolinguistic theory claims that every speech community displays variation, and that variation reflects social structure in an ordered way (Weinreich, Labov, and Herzog 1968: 100). Though some variationist research assumes that roughly the same social factors always underlie variation, more recent authors recognize that the relevant social factors vary between communities, and emphasize the importance of ethnographic grounding for social variables (Schilling-Estes 2013: 8-10, citing Eckert 2000). Nevertheless, sociolinguists working in industrial societies benefit from excellent priors about which aspects of social structure are likely to be relevant to variation. Most documentary fieldworkers do not. At least initially, our priors are poor and biased: poor because the literature on variation in indigenous languages remains small, and biased because outsider researchers bring unconscious assumptions from our own societies to the data. In this study, for example, I immediately saw the importance of gender, a familiar social category, to the variation – but it took me several months of fieldwork to understand the contributions of classificatory kinship, an alien concept. This means that 'doing the local ethnographic footwork' (Rickford 1986: 216) is even more important to sociolinguistics in endangered language settings than elsewhere. Absent ethnography, outsider fieldworkers do not have the evidence or intuitions to form hypotheses about which social factors organize the variation at a field site.

For linguists working on endangered languages more generally, the Northern Máihiki case is important because it adds to a significant body of evidence suggesting that the same factors which cause language endangerment often simultaneously cause dialect mixing. In the Northern Máihiki speech community, both the in-migration of Western Máihiki speakers (the proximate cause of dialect contact) and the endangerment of the variety were due, in part, to a crash in the Máihuna population of the Algodón basin in the early twentieth century. Similarly, Clarke (2009) and Spence (2013) describe North American indigenous communities which experienced simultaneous dialect mixing and language endangerment when people from a large geographic area, speaking a chain of dialects, were compelled to resettle on small reservations with much greater contact between dialect groups. This positions the dialect mixing literature – Kerswill & Williams (2000), Schreier (2003), Kerswill & Trudgill (2005), and Trudgill (2004) on European languages, Clarke (2009) and Spence (2013) on North American ones – as essential reading for language documentarians working in speech communities which have experienced large losses of population or geographic territory.

4. THE LOCUS OF VARIATION. Variation in Northern Máihiki departs from common disciplinary perceptions about variation in two other ways as well. First, while variationist sociolinguistics has tended to focus on phonological and lexical variation, the great majority of inter-speaker variation in Northern Máihiki involves morphosyntax. Second, although the language has no written or formalized standard, some Máihiki speakers engage in Labovian prestige-driven style-shifting between Northern and Western dialect forms. §4.1 discusses the relative prominence of morphosyntactic and phonological variation, and §4.2 addresses style-shifting.

4.1 Morphosyntactic variation. Many of the best-known variationist studies center on phonological and phonetic variation, in line with Neogrammarian tradition, and phonetic variation is the sole concern of the emerging field of sociophonetics. Although variationists have examined morphology, syntax, and pragmatics as well as sound change (Labov 1969, Sankoff & Brown 1976, Tagliamonte & D’Arcy 2007), discussion of variation in endangered languages has, following the lead of the most cited sociolinguistic studies, also focused on phonology. For example, 13 of the 21 chapters in Stanford & Preston (2009) examine phonological variation.

In the Northern Máihiki speech community, on the other hand, morphological and syntactic variation far outstrips phonological variation (cf. Szmrecsanyi & Kortmann 2009). Variation in verbal morphology involves the largest number of variants and the most token-frequent morphs. The speakers listed in Table 1 vary, as shown there, in the form of the affixes marking past tense and 3SG agreement on the main class of verbs and the form of the construction marking past tense and PL or 1SG agreement on a smaller class of extremely frequent verbs. They also exhibit variation in five other aspects of verbal morphology and VP syntax: (a) the perfective construction, (b) temporal overlap (‘while’) clauses, (c) subject control and purpose clauses, and (d-f) the iterative, causative, and terminative verbal affixes. Lesser but still significant variation also occurs in the language’s nominal morphology, particularly its elaborate noun class and noun classification systems, and in the syntax of complex sentences, for instance quotative constructions, copular constructions, and relative clauses. By comparison, the speakers

display variation on only two phonological variables – the reflexes of PM root-medial **h* and root-initial **k^w*.

Máihiki speakers engage in far less metalinguistic talk about morphosyntactic variation than about other forms of variation. Relatedly, the Northern Máihiki speakers who are in close contact with Western dialect speakers have experienced less morphosyntactic change over the lifespan than phonological and lexical change, and no speaker displays different values of morphosyntactic variables between speech styles. These facts indicate that, although morphosyntactic variation in Máihiki is much more extensive than phonological variation and involves more token-frequent elements, speakers nevertheless attach less indexical value to it than to lexical and phonological variation. Like positive ‘anymore’ in some English dialects (Labov 1975: 107, Horn 2014), the morphosyntactic variation is below the limits of metalinguistic awareness. In other speech communities, of course, morphosyntactic variation may be highly indexically salient, as negative concord – characterized by Eckert (2008) as ‘a global resource...in its stereotyped counter-standardness’ – is in English.

4.2 Style-shifting. Variationist literature gives a central place to style-shifting and other forms of intra-speaker variation, and recent ‘third wave’ studies (Bucholtz & Hall 2005, Eckert 2012) – which focus on speakers’ use of linguistic variables to constitute aspects of identity – has placed even greater emphasis on style. Yet work on variation in less-described languages tends to assume that, where a language lacks a written standard or a conventionally formal register, speakers are unlikely to display different variants in response to different linguistic tasks or social situations (Stanford and Preston 2009: 8). The assumption, in slogan form, is that where there is no standard or formal register, there is no style.

For some speakers of Northern Máihiki, though, style is crucial to understanding phonological variation. As mentioned in §3.2.2, the three Northern dialect speakers who have closest contact with Western dialect speakers display Western-like phonological variants in tasks that involve intensive self-monitoring of speech, such as wordlist elicitation, sentence elicitation, and production of extremely careful pronunciations, as in repeating individual words from a text. In tasks which provoke less self-monitoring, such as casual speech and conversation, these speakers exhibit the same Northern phonological variants as the rest of the speech community.

While the range of styles here is compressed, it is clear that these speakers are engaging in the same indexically driven style-shifting observed in variationist sociolinguistic literature. Their use of Western variants in the careful styles cannot represent an attempt to produce a ‘standard’ variety, since Máihiki is rarely written and lacks a meaningful standard. Nor can it reflect accommodation to me, the researcher, since I conducted the elicitation in Spanish. Instead, the Northern Máihiki speakers’ style-shifting likely reflects a metalinguistic evaluation, learned from the in-marrying Western dialect-speaking women, of Northern phonological variants as ‘incorrect’ and Western ones as ‘correct’ (cf. §3.2.2). In this specific case, Western variants are likely evaluated as ‘correct’ because of the greater political power of Western dialect speakers, who are more numerous and were historically less isolated than Northern dialect speakers. More generally, Máihiki speakers’ metalinguistic judgment of the Northern phonological variants illustrates that the availability of ‘correctness’ as an indexical value

does not hinge on the use of different varieties in writing, education, or other formal contexts (cf. Hill & Hill 1986 and Romero 2009 on the indexical value of ‘purity’ in Mesoamerican languages). Even where there is no standard and no formal register, speakers nevertheless command multiple styles and may engage in indexically driven style-shifting.

5. SOCIOLINGUISTICS IN LANGUAGE DOCUMENTATION. To gather the data described in §§3-4 above, I modified my documentary methods in several ways in response to the variation I encountered. My adaptations to variation involved my choice and number of language consultants (§5.1), as well as the kinds of linguistic tasks I did with consultants (§5.2). This section describes these methods in relation to the sociolinguistic and more narrowly documentary goals of the project, and suggests ways in which they could be adapted for other work in documentation of variation.

5.1 Consultants. It is not unusual for descriptive and documentary linguists to work with a relatively small number of language consultants – anecdotes suggest that work with two to five primary consultants is typical. In some cases, endangerment, political issues, or logistics make it impossible to work with a larger number of consultants. In others, fieldworkers prefer to work with fewer consultants so as to facilitate training the consultants for specific linguistic tasks.

Despite these advantages, working with a small number of consultants may conceal both the existence and the structure of variation from the fieldworker. Early in the fieldwork described in §2, influential members of the Northern Máihiki speech community repeatedly pointed out two specific people as good potential consultants for me – Soraida and Adriano (cf. §3.2.2 on these speakers’ variants). Other people, I was told, were not appropriate consultants because they ‘spoke wrong,’ were ‘not from the Algodón,’ or were simply not interested. Although I began working with the two consultants approved by the community, it was clear that there was significant variation both between those two speakers and between them and others whose speech I overheard. Seeking to understand this variation, I began to recruit more consultants, with the goal of recording all 12 known Máihiki speakers born in the Algodón.

As I worked with more consultants, I found that community judgments about speakers’ suitability as consultants rarely agreed with my own. Otilia, labeled by herself and others as ‘speaking wrong,’ turned out to be a fluent, native speaker of both Máihiki and Spanish (a rare combination) and an excellent consultant. The judgment of her Máihiki as ‘wrong’ likely stemmed from her use of certain variants stereotyped as indexing the Eastern dialect, which has very low status in this region. Although findings like this are ubiquitous in sociolinguistics, it is rare to see them cited in language documentation literature – perhaps because documentarians are primed by endangerment to read the statement ‘she speaks wrong’ as ‘she is not a fluent speaker’ rather than as ‘she speaks another variety.’ Likewise, several speakers who others claimed were ‘not from the Algodón,’ such as Trujillo, were in fact born in the Algodón to parents from other settlement areas. Further, many speakers who were characterized as uninterested or too old to work proved to be willing and able to make valuable contributions – a situation familiar to many fieldworkers. If I had followed the community’s initial expectations and not worked with people who ‘spoke wrong’ or were ‘not from here,’ I would have

captured much less of the extent of variation in the speech community, and I doubt that I would have come to understand any of the social patterning of the variation.

This anecdote raises a larger issue for documentation of variation: the differences between linguists' and speech communities' construals of the phrase 'good speaker.' When a fieldworker asks who is a 'good speaker' of the object language, he or she likely means to ask, 'Who is a fluent, native speaker?'. Yet when I asked this question in the Northern Máihiki context, my interlocutors did not answer by naming the fluent, native speakers in the community. Instead, they named a set of people who were fluent speakers and also had high social status for other reasons, such as education or membership in a politically powerful family. My question was construed, effectively, as 'Who is a fluent speaker, and also a person who you feel represents you appropriately to outsiders?' (cf. Evans 2001: 251). Conversely, in language shift situations, fieldworkers may be biased to construe a community member's statement that they speak the object language 'badly' or 'wrong' as meaning that they are a semi-speaker or non-speaker. In the Máihiki context, a person's statement to me that they spoke the language 'wrong' tended to index that they were a fluent speaker with certain low-prestige variants. Semi-speakers occasionally denied that they spoke the language at all, but never described their own speech to me as 'wrong' (again cf. Evans 2001: 261ff.)

I suggest, then, that if documentarians select as consultants only people labeled as 'good speakers,' they will reduce the variation represented in their documentation. In cases where some variants have negative indexical values, this practice will also tend to skew the documentation toward overrepresenting speech which lacks the negatively valued variants – for example, the speech of educated people, men, or particular high-status families – and underrepresenting speech which displays those variants, relative to the true incidence of each class of variants in the speech community (thus the stratified sampling techniques used by variationists). For some documentation projects, it is desirable to embed the community's language and dialect ideologies into the documentation in this way. But for projects which aim to support language maintenance and revitalization activities for people whose heritage variety has the negatively-valued variants, it will be problematic (§6).

5.2 Tasks. While both sociolinguistics and language documentation prioritize the collection of naturalistic or 'free' speech (Tagliamonte, this volume), the two sub-disciplines approach this task differently. Documentation projects often focus on a few highly valued genres of talk, such as traditional narratives or oral history, while sociolinguistic fieldwork prizes 'casual' or 'vernacular' speech and is less concerned with genre (e.g. Schilling-Estes 2013: 98-100, Tagliamonte 2006: 8). Sociolinguistic fieldwork literature assumes that the researcher is a native speaker of a variety mutually intelligible with the object variety (if not of the object variety), and therefore will participate in the discourse which includes the free speech – for example, that he or she will ask questions, contribute new information in response to the speakers, and backchannel. Language documentarians, on the other hand, are not usually native speakers of their object languages. They tend to focus on recording monologic texts, especially traditional narratives, and conversation between consultants, rather than interviews that include the researcher as a focal discourse participant.

While my goals in the fieldwork described in §2 were documentary, I found it essential to adopt two aspects of the ‘sociolinguistic’ approach sketched above in order to record speech showing the full range of variation. First, I decided to record less of the community’s most highly valued genres – myth narrative and song – in order to capture more variation. Although speakers who knew these genres were eager to record them, most speakers did not know them well, and the language of these genres is not representative of spoken Máihiki in general. To obtain free speech samples from every speaker, I requested texts in less stylized genres, such as children’s stories and autobiographical texts (also standby genres for sociolinguistic interviews). These ‘informal’ genres revealed speakers’ style-shifting (§4.2). The autobiographical texts, moreover, proved valuable as sources of both linguistic and social information. The language of these texts provided data for grammatical description and description of variation; the content, information about speakers’ kin relations and personal histories.

Second, I chose to participate in the recorded discourses by commenting on the preceding talk, backchanneling, and occasionally asking questions, in the object language. While participating in the texts became easier as my language ability improved, some strategies, such as asking scripted questions and backchanneling, were feasible from the beginning. These strategies, which are very commonly used by sociolinguists and linguistic anthropologists, encouraged consultants to give longer texts and to produce more grammatically complex stretches of free speech, which were essential for observing morphosyntactic variation (§4.1). My participation also often led consultants to provide genealogical and other background information for my benefit. I note here that it was logistically impossible for me to work with multiple speakers at once in this project. Where possible, recruiting a native speaker to act as an ‘inside interviewer’ would likely be a more effective strategy for obtaining complex free speech. Schilling-Estes (2013: 110-112) offers variationist suggestions for recruiting inside interviewers.

Obtaining free speech from multiple speakers, though, was not sufficient for me to capture the variation. Like most fieldworkers, I check my transcriptions of texts with consultants. This became problematic when the transcription consultant and the author of the text displayed different variants, since – especially for non-salient morphosyntactic variables – the transcription consultant tended to perceive the author as using the same variants as them, even when the author was objectively producing different variants. When the forms in play are phonologically dissimilar, a linguist is likely to notice this kind of mismatch, but when they are similar, the task becomes very difficult. Fieldworkers should therefore be conscious that variation can be erased in the process of transcription, and that the risk of erasing variation increases with one’s degree of reliance on transcriptions made by non-linguists.

6. CONCLUSIONS. For the benefit of other documentary fieldworkers, I have described here some of the ways that encountering the exuberant variation in Northern Máihiki required me to recognize and discard some of my assumptions about variation. The facts of dialect-mixing in this speech community illustrate that the social factors governing variation are far from uniform between speech communities, including between small-scale and non-industrial communities. As a result, some of the social categories which shape variation may not be evident, or even visible, to outsiders (§3). Language and dialect ideology can further conceal variation from outsiders, for instance through style-

shifting (§4.1), in judgments of who is a ‘good speaker’ (§5.1), or in speakers’ inability to perceive certain dimensions of variation (§5.2). And the most important variation may occur beyond the familiar areas of phonetics, phonology, and lexicon (§4.2).

The challenges to studying variation in a language documentation setting are significant, but not insurmountable. They can be addressed, I have suggested, by applying methods from language documentation in concert with the methods of variationist sociolinguistics. To detect variation, it is useful for documentarians to work with consultants from a cross-section of the speech community, as sociolinguists do. Working primarily with people who are described by others as ‘good speakers’ (§5.1), without probing what is meant by ‘good speaker,’ may conceal variation. To apprehend variation that involves style-shifting or morphosyntactic variables, it is important for documentarians to take linguistic samples which are both wide and deep, ideally including spontaneous speech by each consultant in a style that is not highly monitored. Free speech in which the consultant discusses her life history or social position – again, as in sociolinguistic interviews – will be especially useful. Developing the ability to understand the object language, and participating in recordings in a way which displays one’s understanding of the speech, is helpful in collecting ‘casual’ speech and in many other tasks (§5.2).

Documenting variation in the field setting described here had a considerable return on investment for my collaborators and me. In the documentary part of our work, it provided us with a large corpus of texts, from a diverse group of speakers, and improved our internal reconstruction of the language. It also strengthened our relationships with speakers and our understanding of the social dynamics of the Máihiki speech community as a whole.

Documentation of variation is even more important, however, for conservation. Where conservation projects involve any degree of standardization – and even transcription requires some standardization (Ochs 1979) – community members and the linguists who interact with them must know whose variants are being standardized out (cf. Rice and Saxon 2002). Moreover, community-oriented materials, such as learners’ guides and literacy books, may not be acceptable to the community if they do not reflect variation. This holds even where the variation which is removed seems not to be metalinguistically salient. In the Northern Máihiki setting, where metalinguistic talk focused on phonological and lexical variation, speakers still noticed and took exception when community-oriented materials did not reflect their morphosyntactic variants. Linguists can, for some purposes, treat speakers as interchangeable, but those speakers and their descendants may take a different view. The leaders of the Northern Máihiki heritage community made this clear to me: they wanted young people to learn to speak ‘like my sister speaks,’ or ‘like my father spoke,’ not simply to speak Máihiki. For linguists’ work to help heritage communities achieve goals analogous to these, it must include documentation of variation.

REFERENCES

- Austin, Peter K. & Julia Sallabank. 2011. *The Cambridge handbook of endangered languages*. Cambridge: Cambridge University Press.
- Bellier, Irène. 1991. *El temblor y la luna: Ensayo sobre las relaciones entre las mujeres y los hombres mai huna*. Lima: Institut français d'études andines / Abya Yala.
- Bowern, Claire. 2008. *Linguistic fieldwork: A practical guide*. New York: Palgrave Macmillan.
- Campbell, Eric. 2014. *Aspects of the phonology and morphology of Zenzontepec Chatino, a Zapotecan language of Oaxaca, Mexico*. Ph.D. dissertation, University of Texas at Austin.
- Chelliah, Shobhana, & Willem De Reuse. 2011. *The handbook of descriptive linguistic fieldwork*. Dordrecht: Springer.
- Clarke, Sandra. 2009. Social stratification and new dialect formation in a Canadian aboriginal community: Not so different after all?. In James N. Stanford and Dennis Preston (eds.), *Variation in indigenous minority languages*, 109-128. Amsterdam: John Benjamins.
- Eckert, Penelope. 1988. Adolescent social structure and the spread of linguistic change. *Language in Society* 17: 183-207.
- Eckert, Penelope. 2000. *Language variation as social practice: The linguistic construction of identity in Belten High*. Malden: Wiley-Blackwell.
- Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12: 453-476.
- Eckert, Penelope. 2012. Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41: 87-100.
- Evans, Nicholas. 2001. The last speaker is dead – long live the last speaker! In Paul Newman and Martha Ratliff (eds.), *Linguistic fieldwork*, 250-281. Cambridge: Cambridge University Press.
- Gal, Susan. 1979. *Language shift: Social determinants of linguistic change in bilingual Austria*. New York: Academic Press.
- Garde, Murray. 2008. Kun-dangwok: 'Clan lects' and Ausbau in western Arnhem Land. *International Journal of the Sociology of Language* 191: 141-169.
- Givón, T., & Phil Young. 2002. Cooperation and interpersonal manipulation in the society of intimates. *Typological Studies in Language* 48: 23-56.
- Harner, Michael J. 1972. *The Jivaro: People of the sacred waterfalls*. Berkeley: University of California Press.
- Hayden, Brian, & Aubrey Cannon. 1982. The corporate group as an archaeological unit. *Journal of Anthropological Archaeology* 1: 132-158.
- Hill, Jane H., & Kenneth C. Hill. 1986. *Speaking Mexicano: Dynamics of syncretic language in central Mexico*. Tucson: University of Arizona Press.
- Horn, Laurence R. 2014. Afterword: Microvariation in syntax and beyond. In Raffaella Zanuttini & Laurence R. Horn, eds., *Micro-syntactic variation in North American English*, 324-347. Oxford: Oxford University Press.
- Hugh-Jones, Christine. 1979. *From the Milk River: Spatial and temporal processes in the Northwest Amazon*. Cambridge: Cambridge University Press.
- Jackson, Jean. 1983. *The fish people: Linguistic exogamy and Tukanoan identity in*

- Northwest Amazonia. Cambridge: Cambridge University Press.
- Johnson, Alan. 2003. *Families of the forest: The Matsigenka Indians of the Peruvian Amazon*. Berkeley: University of California Press.
- Kerswill, Paul & Peter Trudgill. 2005. The birth of new dialects. In P. Auer, F. Hinskens, and P. Kerswill (eds.), *Dialect change: Convergence and divergence in European languages*, 196-220. Cambridge: Cambridge University Press.
- Kerswill, Paul & Ann Williams. 2000. Creating a new town koine: Children and language change in Milton Keynes. *Language in Society* 29: 65-115.
- Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45: 715-762.
- Labov, William. 1975. Empirical foundations of linguistic theory. In Robert Austerlitz, ed., *The scope of American linguistics*, 77-133. Lisse: The Peter de Ridder Press.
- Labov, William. 1981. Resolving the Neogrammarian controversy. *Language* 57: 267-308.
- Labov, William. 2001. *Principles of linguistic change, Vol. 2: Social factors*. Malden: Wiley.
- Lave, Jean & Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation*. Palo Alto, CA: Institute for Research on Learning.
- Lippi-Green, Rosina L. 1989. Social network integration and language change in progress in a rural Alpine village. *Language in Society* 18: 213-234.
- Meyerhoff, Miriam. 2015. Turning variation on its head: Analysing subject prefixes in Nkep (Vanuatu) for language documentation. *Asia-Pacific Language Variation* 1: 78-108.
- Milroy, Lesley. 1980. *Language and social networks*. Oxford: Blackwell.
- Milroy, James and Lesley Milroy. 1985. Linguistic change, social network, and speaker innovation. *Journal of Linguistics* 21: 339-384.
- Ochs, Elinor. 1979. Transcription as theory. In Elinor Ochs and Bambi B. Schieffelin (eds.), *Developmental pragmatics*, 43-72. New York: Academic Press.
- Rice, Keren. 1989. *A grammar of Slave*. Berlin: De Gruyter.
- Rice, Keren. 2006. A typology of good grammars. *Studies in Language* 30: 385-415.
- Rice, Keren, & Leslie Saxon. 2002. Issues of standardization and community in Aboriginal language lexicography. In William Frawley, Kenneth C. Hill, & Pamela Munro, eds., *Making dictionaries: Preserving indigenous languages of the Americas*, 125-154. Berkeley: University of California Press.
- Rickford, John R. 1986. The need for new approaches to social class analysis in sociolinguistics. *Language and Communication* 6: 215-221.
- Romero, Sergio. 2009. Phonological markedness, regional identity, and sex in Mayan: the fricativization of intervocalic /l/ in K'iche'. In James N. Stanford and Dennis Preston (eds.), *Variation in indigenous minority languages*, 281-297. Amsterdam: John Benjamins.
- Samarin, W. J. 1967. *Field linguistics*. New York: Rinehart & Winston.
- Sankoff, Gillian, and Penelope Brown. 1976. The origins of syntax in discourse: A case study of Tok Pisin relatives. *Language* 52: 631-666.
- Schilling-Estes, Natalie. 2013. *Sociolinguistic fieldwork*. Cambridge: Cambridge University Press.
- Schreier, Daniel. 2003. *Isolation and language change: Contemporary and*

- sociohistorical evidence from Tristan da Cunha English*. New York: Palgrave Macmillan.
- Sorensen, Arthur. 1967. Multilingualism in the Northwest Amazon. *American Anthropologist* 69: 670-684.
- Spence, Justin. 2013. *Language change, contact, and koineization in Pacific Coast Athabaskan*: University of California Berkeley dissertation.
- Stanford, James N. 2008. Child dialect acquisition: new perspectives on parent/peer influence. *Journal of Sociolinguistics* 12: 567-596.
- Stanford, James N. 2009. Clan as a sociolinguistic variable: three approaches to Sui clans. In James N. Stanford & Dennis Preston (eds.), *Variation in indigenous minority languages*, 463-484. Amsterdam: John Benjamins.
- Stanford, James N. & Dennis Preston. 2009. The lure of a distant horizon: variation in indigenous minority languages. Introduction to James N. Stanford & Dennis Preston (eds.), *Variation in indigenous minority languages*. Amsterdam: John Benjamins.
- Stanford, James N. & Yanhong Pan. 2013. The sociolinguistics of exogamy: Dialect acquisition in a Zhuang village. *Journal of Sociolinguistics* 17: 573-607.
- Szmrecsanyi, Benedikt, & Bernd Kortmann. 2009. The morphosyntax of varieties of English worldwide: A quantitative perspective. *Lingua* 119: 1643-1663.
- Tagliamonte, Sali. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.
- Tagliamonte, Sali. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Malden: Wiley-Blackwell.
- Tagliamonte, Sali & Alexandra D'Arcy. 2007. Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change* 19: 199-217.
- Trudgill, Peter. 1974. *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Trudgill, Peter, Elizabeth Gordon, Gillian Lewis & Margaret MacLagan. 2000. Determinism in new-dialect formation and the genesis of New Zealand English. *Journal of Linguistics* 36: 299-318.
- Trudgill, Peter. 2004. *New-dialect formation: The inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.
- Trudgill, Peter. 2011. *Sociolinguistic typology: The social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Vaux, Bert & Justin Cooper. 1999. *Introduction to linguistic field methods*. Munich: Lincom Europa.
- Velie, Daniel. 1975. *Bosquejo de la fonología y gramática del idioma orejón (coto)*. Lima: SIL.
- Velie, Daniel & Virginia Velie. 1981. *Vocabulario orejón*. Lima: SIL and Ministerio de Educación.
- Weinrich, Uriel, William Labov, & Marvin I. Herzog. 1968. Empirical foundations for a theory of language change. In W.P. Lehmann and Yakov Malkiel, eds., *Directions for historical linguistics*, 95-195. Austin: University of Texas Press.